Why (authors believe that) reviewers stress limiting aspects of manuscripts: The SLAM effect in peer review.

van Lange, P.A.M.

published in
Journal of Applied Social Psychology
1999

DOI (link to publisher)
10.1111/j.1559-1816.1999.tb00125.x

document version
Publisher's PDF, also known as Version of record

Why Authors Believe That Reviewers Stress Limiting Aspects of Manuscripts: The SLAM Effect in Peer Review[1]

PAUL A. M. VAN LANGE[2]
Department of Social Psychology
Free University
Amsterdam, The Netherlands

This manuscript describes a preliminary study examining judgments of authors and reviewers regarding manuscripts that have been either accepted or rejected for publication. Consistent with hypotheses, results reveal that participants believe that their own manuscripts are superior to others' manuscripts in terms of general, theoretical, and methodological quality. Relevant to the presumed tendency among reviewers to *stress limiting aspects of manuscripts* (SLAM), reviewers exhibited greater agreement with editorial decisions favoring rejection, relative to those favoring acceptance. These findings suggest that authors' beliefs in reviewers' tendencies to SLAM can be partially understood in terms of authors' unrealistically favorable and optimistic beliefs regarding their manuscripts and in reviewers' actual tendencies to be quite critical—at least more critical than editors.

Virtually all sciences rely on the peer-review system, a practice that has been discussed by various scientists. While there is a good deal of agreement among scientists of different disciplines regarding the overall utility of the peer-review system, few (if any) believe that the peer-review system is without limitations (e.g., Laband & Piette, 1994; Peters & Ceci, 1985).

Recently, Epstein (1995) fostered further debate regarding the peer-review system, noting that reviewers tend to place particular emphasis on (often seemingly correctable) imperfections and limitations, rather than on the innovative aspects of a manuscript. He summarized a discussion with fellow scientists (several of whom were associate editors who had many publication credits in APA

[1]This research was supported in part by a grant from the Netherlands Organization for Scientific Research (NWO; R-57-178). I thank the organizers of two conferences, Chuck Stangor and Wolfgang Wagner, for their help in conducting this research; Wilma Otten and two anonymous reviewers for their constructive comments on an earlier draft of this manuscript; and all of the social psychologists who generously participated in this research.

[2]Correspondence concerning this article should be addressed to Paul Van Lange, Department of Social Psychology, Free University at Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. e-mail: pam.van.Lange@psy.vu.nl.

Journal of Applied Social Psychology, 1999, **29**, 12, pp. 2550-2566.

journals and who did a considerable amount of reviewing themselves) by noting that "I was impressed by the widespread discontent with the journal review process in this select group. Some of the terms people used to describe the reviews were 'arbitrary,' 'biased,' 'self-serving,' 'irresponsible,' and 'arrogant'" (p. 884).

Epstein (1995) further illustrated such "biases" among reviewers by suggesting that reviewers frequently exhibit an "I gotcha" mentality, a frame of mind in which reviewers are tempted to recommend rejection on the basis of minor limitations, even in case of an otherwise excellent manuscript (for similar observations, see Rabinovich, 1996). Are reviewers really that bad? Does this belief in what I refer to as reviewers' tendencies to SLAM (an acronym for *stressing limiting aspects of manuscripts*) reflect an unbiased, accurate judgment? Or might this belief be colored by authors' tendencies to interpret reviews (and reviewers' inclinations) in ways so as to maintain a favorable view of one's own manuscript?[3]

While the belief in reviewers' tendencies to SLAM seems to be shared by many researchers, Levenson (1996) pointed out that it is common for authors to engage in rationalization after reading reviews of their manuscripts, particularly those reviews that favor rejection rather than acceptance. He referred to this phenomenon as the *sour-grapes hypothesis*, stressing the notion that reviewers' tendencies to SLAM are, to some degree, constructed beliefs. Indeed, it is unlikely that reviewers themselves describe their reviews in terms of SLAM. It is even less likely that reviewers themselves describe their reviews as "'arbitrary,' 'biased,' 'self-serving,' 'irresponsible,' and 'arrogant'" (Epstein, 1995, p. 884), as Epstein's fellow scientists characterized reviews. Instead, it is more likely that reviewers believe that their (negative) appraisals are entirely justified, in light of what are believed to be serious limitations underlying the work they reviewed (e.g., "The theory was poorly developed," "The studies were so poorly designed"; cf. Levenson, 1996). Thus, it is possible that the belief in reviewers' tendencies to SLAM represents an objective reality, a constructed reality, or both.

A Preliminary Study of Author and Reviewer Judgments

In light of the preceding discussion—and given that the accumulation of knowledge relies to a significant degree on the peer-review system—it becomes important to examine the experiences of authors *and* reviewers. The present

[3]Unfortunately, the acronym *SLAM* may have some specific connotations that I do not intend to convey. For example, one connotation might be that reviewers are merely interested in providing authors with negative feedback. Of course, I do not use this acronym because of such harsh connotations, which in my view provide a very poor and inaccurate description of reviewers' motivations. Nevertheless, I believe that this acronym captures the global meaning that I wish to convey: (authors believing that) reviewers stress limitations more than they do strengths. I thank an anonymous reviewer for suggesting this acronym.

manuscript describes a preliminary study, designed to enhance our understanding of (a) how authors judge the quality of their own manuscripts that have been either accepted or rejected for publication, (b) how such judgments relate to judgments of others' manuscripts that have been either accepted or rejected for publication, and (c) the extent to which authors and reviewers agree or disagree with editorial decisions favoring acceptance and rejection. Specifically, this research asks several social psychologists to think of the most recent manuscript that they submitted (vs. reviewed) and that has been accepted (vs. rejected) for publication. Thereafter, they are asked to judge the general quality, theoretical quality, and methodological quality of the manuscript and to indicate how much they agree with the editorial decision.

Our framework for understanding experiences of authors and reviewers regarding the peer-review system proposes that evaluations of own and others' work to some extent are socially defined and colored by relatively stable beliefs of superiority, assuming that one's own work is better than and not as bad as others' work. This proposition may be derived from classic and contemporary assumptions underlying theories of social comparison and self–other judgment, which emphasize the social and self-enhancing nature of judgments, particularly in contexts in which perfectly objective standards for judgment are not available (cf. Festinger, 1954; Suls & Wills, 1991; Taylor & Brown, 1988). Accordingly, the perceived quality of own manuscripts may be partially affected by the good and bad features believed to be associated with others' manuscripts (e.g., "This paper is actually quite good, when I look at what is being published these days"). Similarly, the perceived quality of others' manuscripts may be partially affected by the good and bad features believed to be associated with own manuscripts (e.g., "Compared to my own work, this work does not seem to be all that novel").

Congruent with these theoretical approaches, there is considerable evidence that judgments of others are affected by beliefs about the self (and vice versa), and that individuals tend to regard themselves as being better than average on several competence-related attributes (Suls & Wills, 1991; Taylor & Brown, 1988; see also Allison, Messick, & Goethals, 1989; Van Lange & Rusbult, 1995). This evidence suggests that people have developed relatively stable views of the self, frequently referred to as *positive self-schemas* (Taylor & Brown, 1988), which may be used to interpret, filter, and color new self-relevant information. In part, such information processing is guided by *self-enhancement*, the motivation to elevate the positivity of one's self-conceptions and to protect one's self-concepts from negative evaluation. Also, such information processing might be guided by self-verification, the motivation to maintain consistency between self-conceptions and new self-relevant information (Swann, 1983, 1990; see also Sedikides & Strube, 1997). Indeed, if people already hold beliefs of self–other superiority, self-enhancement and self-verification are highly complementary mechanisms.

Importantly, given that beliefs of superiority may to some degree serve as an anchor for evaluating others' work—and given that it is not easy to compete with such standards—we as reviewers are likely to take a fairly critical approach to the work of others, thereby stressing the limitations rather than the strengths of the manuscripts we review. For example, consciously or unconsciously, we as reviewers may sometimes be tempted to recommend rejection on the basis of the (implicit but often disputable) belief that the limitations of the work we review are (far) more serious than the limitations that others have stressed regarding our own work.[4] It is also interesting to note that by communicating experiences regarding "they as reviewers" and "we as reviewers," we are likely to confirm our beliefs of superiority.

Most of our colleagues can probably relate to the feeling that reviewers of our work (particularly work that has been rejected) can be somewhat biased or shortsighted and to the feeling that we as reviewers every now and then devote time to very poor manuscripts. Accordingly, the belief in reviewers' tendencies to SLAM is likely to represent an objective reality (i.e., on average, reviewers do place particular emphasis on limitations) and a constructed reality (i.e., authors do hold unrealistically favorable beliefs about their own research).

Hypotheses: Beliefs of Superiority and Beliefs in Reviewers' Tendencies to SLAM

On the basis of the preceding lines of reasoning, three hypotheses are advanced. First, it is predicted that social psychologists will evaluate their own manuscripts more favorably than those of others in terms of general quality, theoretical quality, as well as methodological quality. Second, perceptions of superiority may also to some extent affect the degree to which authors and reviewers agree or disagree with editorial decisions. Accordingly, it is predicted that for manuscripts being accepted, levels of agreement with the editorial decision will be greater for own manuscripts than for others' manuscripts. Conversely, for manuscripts being rejected, levels of agreement with the editorial decision should be lower for own manuscripts than for others' manuscripts.

[4]However, evaluations need not always be socially defined or colored by beliefs of superiority. For example, sometimes it is not difficult to evaluate a manuscript, in that its quality is unequivocally excellent or unequivocally poor. Under such circumstances, our evaluations do not tend to be socially defined, although such review experiences may activate further reasoning that can be understood in terms of beliefs of superiority. For example, an unequivocally poor manuscript may help researchers to maintain a relatively favorable view of their own work (e.g., "Even my worst manuscript is better"; cf. downward comparison, Wills, 1991). And, an unequivocally excellent manuscript may instigate somewhat exaggerated beliefs about the author of this manuscript ("This researcher must be a genius"; cf. Alicke, LoSchiavo, Zerbst, & Zhang, 1997), so that reviewers are still able to believe that their own work is quite good. Thus, evaluations may be socially defined and affected by beliefs of superiority, especially in the important gray area, ranging from not particularly good to quite good manuscripts.

Finally, this research seeks to provide evidence relevant to notion that the belief in reviewers' tendencies to SLAM to some degree reflects an objective reality. However, it is not easy to provide direct evidence in support of this notion. How can one judge that reviewers are actually too critical? How can one judge that reviewers "overstress" limitations? This issue is approached by examining reviewers' levels of agreement regarding accept versus reject decisions taken by the editor, thus employing the editorial decision as a benchmark for assessing reviewers' tendencies to favor rejection rather than acceptance. Granted, this benchmark is indirect, in that the relationship between perceptions of limitations and recommendations for rejection (while pronounced) is unlikely to be perfect. At the same time, this benchmark does take account of the low acceptance rates, a standard which is likely to inspire some tendency to SLAM among reviewers and editors (cf. Reis & Stiller, 1992). Thus, based on the assumption that the belief in SLAM among reviewers to some degree reflects an objective reality, it is predicted that reviewers will tend to agree more with editorial decisions favoring rejection than with editorial decisions favoring acceptance.

Method

Participants

Participants were recruited at two major conferences of social psychology, including a conference (in Washington, DC, in September 1995) jointly organized by the Society of Experimental Social Psychology (SESP) and the European Association of Experimental Social Psychology (EAESP), and a conference held in Gmunden (Austria, in July 1996) organized by the EAESP. At both conferences, questionnaires were distributed, which participants could complete either during the conference (and drop in a designated box) or after the conference (and mail to the university). Three weeks after each conference, a total of 54 questionnaires was returned (26 questionnaires for the SESP/EAESP joint meeting; 28 questionnaires for the EAESP conference).

In light of the large number of social psychologists who attended these conferences (i.e., both conferences were attended by at least 300 social psychologists), the response rate is, of course, not ideal (even when one takes into account the fact that a fair number attended both conferences). However, an impressive response rate was not anticipated, for two reasons. First, participants of conferences receive a fair amount of material and tend to be quite busy during conferences. Second, many participants indicated that they wanted to participate, but could not do so because they had not yet served as reviewers for a peer-reviewed journal or had not yet published papers. Also, given that participants from two different conferences were recruited, I was able to see whether there might be

any substantial differences as a result of sample or selection differences. Examination of possible differences between the two samples for both principal dependent measures (i.e., quality judgments and level of agreement with the editorial decision) reveals no evidence for any significant main or interaction effects involving "type of conference." Thus, there is some (albeit indirect) support for the generality of our findings.

Of the 54 questionnaires returned, 6 could not be used. That is, 1 questionnaire contained several missing values, 2 questionnaires indicated that these participants had not yet submitted papers that were accepted for publication, and 3 questionnaires indicated that these participants had not yet served as reviewers for a peer-reviewed journal. The remaining sample of 48 participants (35 men, 12 women, 1 failed to indicate gender; M age = 42 years) consisted of social psychologists working in various countries, including Austria, Australia, Belgium, Canada, England, France, Germany, Greece, Italy, Israel, the Netherlands, Sweden, Portugal, the United States, New Zealand, and Wales. The sample consisted of 42 professors (17 full professors, 16 associate professors, 9 assistant professors), 3 participants were post-doctoral research fellows, and 3 were graduate students. All of the participants had submitted manuscripts to social psychological or related journals and had served as reviewers for peer-reviewed journals.

Experimental Design

This study employs a 2 × 2 (Author: Own Manuscript vs. Others' Manuscript × Editorial Decision: Accept vs. Reject) between-participants design. The dependent measures include (a) judgments of general quality, theoretical quality, and methodological quality of a submitted paper; and (b) level of agreement with the editorial decision.

Procedure

The first page of the questionnaire, entitled "The Social Psychology of Social Psychologists," describes (a) the purpose of the research in very general terms (i.e., examining experiences of social psychologists), (b) anonymity of participants' responses, and (c) debriefing procedures. To ensure anonymity and to provide the possibility for debriefing, I asked participants to return stickers with their names and addresses in a different box at the registration desk. At the SESP/EAESP joint meeting, the questionnaire consisted of different parts, examining issues such as descriptions of interaction situations at the conference, and evaluations and judgments of person talk (e.g., gossip). Because these topics were not central to the current research, they will not be discussed further. At the EAESP conference, the questionnaire examined only judgments of manuscripts.

As noted earlier, there were four conditions, based on the factorial crossing of author (own vs. others' manuscript) and editorial decision (accept vs. reject). In the own-manuscript–accept condition, the instructions read, "For the questions below we would like to ask you to think about the empirical paper that you most recently *submitted as primary author to* a peer-reviewed journal and that has been *accepted* for publication." In the others'-manuscript–accept condition, the phrase *submitted as primary author to* was replaced with *reviewed for*. In the rejection conditions (i.e., own-manuscript–reject and others'-manuscript–reject condition), the word *accepted* was replaced with *rejected*.

In total, eight questions were asked. These questions focused on evaluations of a paper in its first submitted version so as to yield comparable judgment situations for the two author conditions (own vs. others' manuscript). That is, reviewers typically do not have access to information relevant to the quality of a subsequent revision, whereas authors do. A first submitted version tends to differ in a number of aspects from a revision of this first draft. For example, relative to the first submitted paper, a revision of that paper is associated with improved quality (or at least the perception thereof) and a stronger awareness of the strengths and limitations of the study or studies described. Thus, to avoid such asymmetries in evaluations of own and others' manuscripts, I examined evaluations of the first submitted version.

Five questions focused on judgments of quality. First, this research assessed ratings of general quality ("How would you rate the overall quality of the paper in its first submitted version?") on a 7-point scale ranging from 1 (*very poor*) to 7 (*excellent*). Second, two questions assessed ratings of theoretical quality of the paper ("How would you rate the theoretical aspects of the paper in its first submitted version?" was rated on the same 7-point scale as the one for general quality; "To what extent were there theoretical ambiguities in the paper [e.g., logic was not entirely clear; a theoretically relevant link was missing]?" was rated on a 7-point scale ranging from 1 [*there were many ambiguities*] to 7 [*there were no ambiguities*]). The correlation between these two ratings was fairly high, $r(48) = .69$, so the average of these two ratings was used in subsequent analyses. Third, using the same wording and scales as for the assessment of theoretical quality, this research assessed two ratings of the methodological quality of the paper (the illustration for methodological ambiguities was "incomplete information regarding procedure or analyses"). The correlation between these two ratings was fairly high, $r(48) = .67$, so the average of these two ratings was used in subsequent analyses. A final question assessed degree of agreement with the editorial decision by asking participants to what extent they agreed or disagreed with the editorial decision on a 7-point scale ranging from 1 (*I disagree completely*) to 7 (*I agree completely*).

I also asked each participant whether he or she would be willing to share the name of the journal that had accepted or rejected the manuscript. The most

frequently listed journals were *Journal of Personality and Social Psychology* (*JPSP*; 14 times) and *Personality and Social Psychology Bulletin* (*PSPB*; 16 times). In a highly qualitative manner, I explored whether these two titles were about equally distributed over the four conditions. It appeared that *JPSP* was listed at least twice in each condition, and *PSPB* was listed at least once in all four conditions. Although this evidence is clearly impressionistic, it did not seem to be the case that the quality of the journals varied substantially across the four conditions.

Results

The analyses proceeded in two stages. To begin with, the hypotheses regarding quality judgments and level of agreement with the editorial decision were tested using a series of 2 × 2 (Author × Editorial Decision) ANOVAs. However, as the reader will note, the assumption of homogeneity of variance was violated (i.e., standard deviations varied significantly across differing cells). Even when the data are transformed (i.e., square-root transformation or log transformation), the standard deviations among cells were significantly different. Although parametric tests such as ANOVA tend to be fairly robust to violations of the assumption of homogeneity, nonparametric tests are arguably more appropriate. The results of the ANOVA will nevertheless be reported because I regard the reality comprising differences in means and standard deviations to be substantially meaningful. Moreover, ANOVAs focusing on one of the two groups of participants (i.e., joint SESP/EAESP meeting participants and EAESP conference participants) yielded identical findings, but only occasionally revealed a significant violation of the assumption of homogeneity. After describing the results of ANOVAs, I report the results of a series of loglinear analyses, for which this assumption is irrelevant, thus complementing the ANOVA with statistically more appropriate tests of my hypotheses.

Are Own Manuscripts Perceived to Be Superior to the Manuscripts We Review?

I submitted the three ratings of quality (general, theoretical, and methodological) to a 2 × 2 (Author: Own vs. Others' Manuscripts × Editorial Decision: Accept vs. Reject) MANOVA. This analysis reveals a significant multivariate main Effect for author, $F(3, 42) = 5.26$, $p < .005$, which at the univariate level reveals a significant effect for general quality, $F(1, 44) = 14.18$, $p < .001$; theoretical quality, $F(1, 44) = 11.52$, $p < .001$; and methodological quality, $F(1, 44) = 10.88$, $p < .005$. As can be seen in Figure 1, results illustrate that, relative to others' manuscripts, own manuscripts were rated to be superior in terms of general quality (*M*s = 3.80 vs. 5.21, respective *SD*s = 1.64 and 0.69), theoretical quality

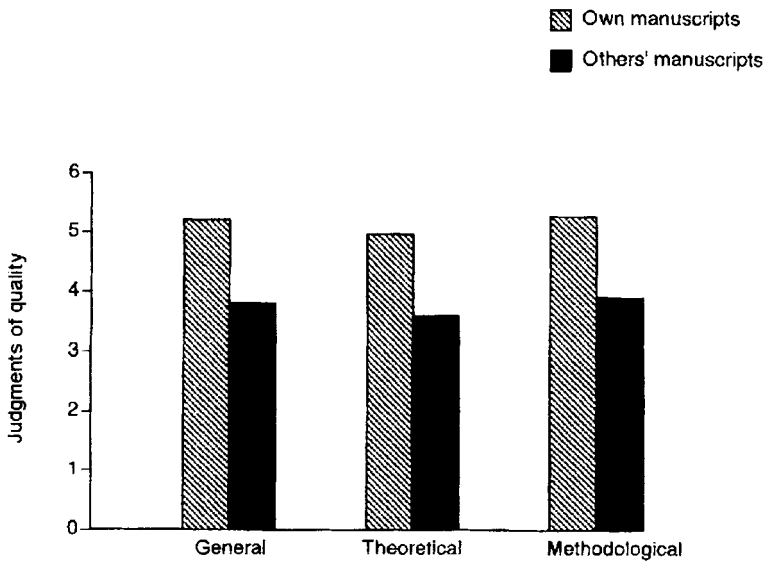

Figure 1. Judgments of quality regarding own and others' manuscripts.

(*M*s = 3.60 vs. 4.96, respective *SD*s = 1.63 and 0.85), and methodological quality (*M*s = 3.90 vs. 5.27, respective *SD*s = 1.73 and 0.78).

Not surprisingly, the analysis also reveals a significant multivariate main effect for editorial decision, $F(3, 42) = 2.95$, $p < .05$, which reflects a main effect for general quality, $F(1, 44) = 5.61$, $p < .05$; theoretical quality, $F(1, 44) = 5.15$, $p < .05$; and methodological quality, $F(1, 44) = 8.11$, $p < .01$. Relative to manuscripts that were rejected, manuscripts that were accepted for publication were rated to be superior in terms of general quality (*M*s = 4.13 vs. 5.08, respective *SD*s = 1.42 and 1.15), theoretical quality (*M*s = 3.89 vs. 4.86, respective *SD*s = 1.41 and 1.25), and methodological quality (*M*s = 4.09 vs. 5.26, respective *SD*s = 1.34 and 1.28). No further effects were significant.

Could it be that we believe that our own rejected manuscripts are not inferior to the manuscripts that we have reviewed and that have been accepted for publication? Specific contrasts reveal that own rejected manuscripts were believed to be at least as good as were others' manuscripts that have been accepted. None of the contrasts were significant, and the means for general quality (*M*s = 5.09 vs. 4.63, respective *SD*s = 0.71 vs. 1.77), theoretical quality (*M*s = 4.77 vs. 4.38, respective *SD*s = 0.90 vs. 1.85), and methodological quality (*M*s = 4.91 vs. 4.75, respective *SD*s = 0.66 vs. 1.95) indicate that we certainly do not regard our own rejected manuscripts as inferior to others' accepted manuscripts.

As noted earlier, the standard deviations among cells were substantially different, as revealed by significant Bartlett-Box tests for judgments of general

quality, $F(3, 2903) = 4.48, p < .01$; theoretical quality, $F(3, 2903) = 2.83, p < .05$; and methodological quality, $F(3, 2903) = 4.72, p < .005$. For each judgment, the standard deviations were smaller for judgments of own manuscripts (i.e., *SD*s ranged from 0.66 to 0.91 across both conditions) than for judgments of others' manuscripts (i.e., *SD*s ranged from 1.31 to 1.85 across both conditions). Thus, irrespective of whether own manuscripts were accepted or rejected, almost all of the participants believed that their work was at least quite good (i.e., none of the participants' ratings of general quality were smaller than the midpoint of the scale). In contrast, there was considerable variation in quality judgments regarding others' work (i.e., 10 ratings were below the midpoint and 8 were above the midpoint of the scale).

To examine the three judgments of quality in a statistically appropriate manner, three 2 × 2 × 2 (Quality: High vs. Low × Author × Editorial Decision) loglinear analyses were conducted. In each analysis, judgments of quality (i.e., general, theoretical, and methodological quality) were dichotomized by comparing ratings of 4 or lower (i.e., low quality) with ratings that exceeded this midpoint of the scale (i.e., high quality). Paralleling our earlier main effects for author, all three analyses reveal a significant interaction of author and quality: general quality, $\chi^2(1, N = 48) = 9.47, p < .005$; theoretical quality, $\chi^2(1, N = 48) = 12.10, p < .001$; and methodological quality, $\chi^2(1, N = 48) = 7.67, p < .01$.

Results reveal that 24 of 28 participants (86%) considered their own work of high general quality, whereas only 8 of 20 participants (40%) considered others' work of high general quality (40%). Similar effects were found for judgments of theoretical and methodological quality. A majority (23 of 28 participants, or 82%) considered their own work of high theoretical quality, whereas only 6 of 20 participants (30%) considered others' work of high theoretical quality. Similarly, 24 of 28 participants (86%) considered their own work of high methodological quality, whereas only 9 of 20 participants (45%) considered others' work of high methodological quality. None of the analyses reveal any other significant effects, except for a main effect of quality in the analyses of general quality, $\chi^2(1, N = 48) = 5.44, p < .05$, and methodological quality, $\chi^2(1, N = 48) = 6.92, p < .05$, indicating that a small majority regarded the overall quality to be high (i.e., a small majority exhibited above-midpoint ratings for general quality [67%] and methodological quality [69%]). Thus, these analyses, too, are consistent with our hypothesis stating that participants believe that their own manuscripts are superior to others' manuscripts in terms of general quality, theoretical quality, and methodological quality.

When Do We Agree and When Do We Disagree With Editorial Decisions?

Levels of agreement with editorial decisions were submitted to a 2 × 2 (Author: Self vs. Other × Editorial Decision: Accept vs. Reject) ANOVA. This

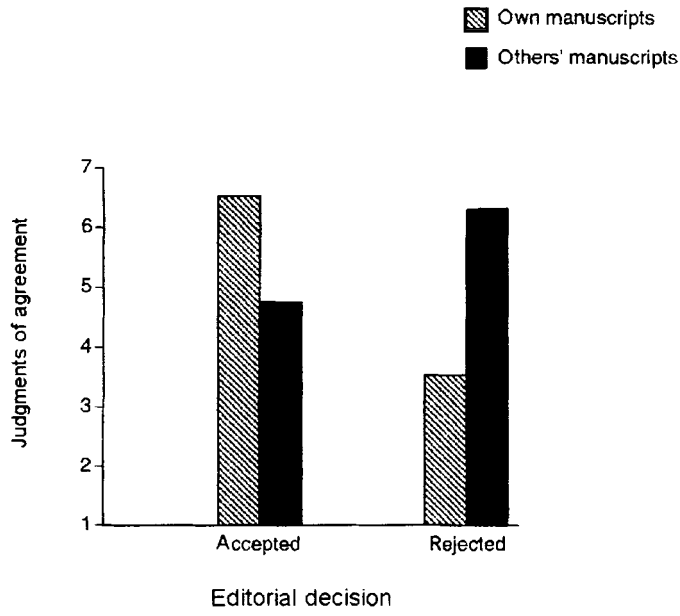

Figure 2. Levels of agreement with editorial decisions favoring acceptance or rejection of own and others' manuscripts.

analysis reveals an interaction of author and editorial decision, $F(1, 44) = 30.84$, $p < .001$, and no significant main effects. As can be seen in Figure 2, simple effects reveal that for manuscripts being accepted, levels of agreement were greater for own manuscripts ($M = 6.53$, $SD = 0.62$) than for others' manuscripts ($M = 4.75$, $SD = 2.31$), $F(1, 44) = 12.54$, $p < .001$. Conversely, for manuscripts being rejected, levels of agreement were lower for own manuscripts ($M = 3.55$, $SD = 1.69$) than for others' manuscripts ($M = 6.33$, $SD = 0.99$), $F(1, 44) = 22.86$, $p < .001$.

Do we as reviewers exhibit lower levels of agreement with editorial decisions favoring acceptance than those favoring rejection? Consistent with hypotheses, simple effects reveal, for others' manuscripts, greater levels of agreement for rejected manuscripts ($M = 6.33$, $SD = 0.99$) than for accepted manuscripts ($M = 4.75$, $SD = 2.31$), $F(1, 44) = 6.89$, $p = .01$. Of lesser relevance (although congruent with our hypotheses), simple effects for own manuscripts reveal greater levels of agreement with editorial decisions favoring acceptance ($M = 6.53$, $SD = 0.62$) than those favoring rejection ($M = 3.55$, $SD = 1.70$), $F(1, 44) = 29.54$, $p < .001$.

As was the case for judgments of quality, the variances among cells were substantially different, as revealed by a significant Bartlett-Box test, $F(3, 2903) = 6.80$, $p < .001$. Unlike the pattern for quality judgments, for level of agreement, variances were substantially lower when own manuscripts were accepted or

when others' manuscripts were rejected (both *SD*s < 1, as noted earlier) than when own manuscripts were rejected or when others' manuscripts were accepted (both *SD*s > 1.69). Thus, for own accepted manuscripts and for others' rejected manuscripts, almost all of the participants exhibited very high levels of agreement with the editorial decision.

As in the analyses of quality judgements, the association among variables was analyzed using a 2 × 2 × 2 (Agreement: High vs. Low × Author × Editorial Decision) loglinear analysis. Given that the overall levels of agreement were fairly high, I compared ratings of 5 or lower (i.e., low agreement) with ratings that exceeded 5 (i.e., high quality), resulting in a more equal distribution than the one determined by the midpoint of the scale (even though 65% exhibited ratings greater than 5). Paralleling our earlier interaction of author and editorial decision, this analysis reveals a significant three-way interaction of agreement, author, and quality, $\chi^2(1, N = 48) = 19.82$, $p < .001$. For own manuscripts, a large majority (16 of 17 participants, 94%) indicated high agreement with acceptance, whereas for others' manuscripts, a large majority (10 of 12 participants, 83%) indicated high agreement with rejection. Indeed, a specific follow-up analysis focusing on manuscripts that the participants had reviewed (i.e., others' manuscripts) reveals a significant association between agreement (high vs. low) and editorial decision, $\chi^2(1, N = 20) = 4.43$, $p < .05$. Thus, these analyses, too, are consistent with our hypotheses, stating that (a) participants exhibit greater levels of agreement with acceptance of own versus others' manuscripts, and lower levels of agreement with rejection of own versus others' manuscripts; and (b) for others' manuscripts, participants exhibit greater levels of agreement with editorial decisions favoring rejection than those favoring acceptance.

Links Between Perceived Quality and Agreement

Congruent with the assumption that quality judgments affect levels of agreement with editorial decisions, results reveal that for accepted manuscripts, greater levels of perceived general quality, theoretical quality, and methodological quality were associated with greater levels of agreement (respective *r*s = .46, .34, and .64, all *p*s < .10). Conversely, for rejected manuscripts, greater levels of perceived general quality, theoretical quality, and methodological quality were associated with lower levels of agreement (respective *r*s = -.57, -.60, and -.60, all *p*s < .01). Thus, agreement with editorial decisions seems to be partially based on judgments of theoretical as well as methodological quality.

Discussion

The current study provides good evidence in support of the claim that social psychologists tend to hold superior beliefs about their own manuscripts,

ascribing greater quality (i.e., general quality, theoretical quality, and methodological quality) to the manuscripts that they have submitted than to the manuscripts they have reviewed. Moreover, levels of agreement with an editorial decision were greater for own (vs. others') manuscripts that have been accepted for publication; conversely, levels of agreement with an editorial decision were greater for others' (vs. own) manuscripts that have been rejected for publication. And in the role of reviewers, participants exhibited greater agreement with editorial decisions favoring rejection rather than acceptance, suggesting that reviewers are more critical than are editors.

Generally, these findings are consistent with the assumption that judgments of own and others' work to some extent are socially defined and colored by relatively stable beliefs of superiority; processes that can be understood in terms of self-enhancement and self-verification. When submitting manuscripts, many of us would seem to be somewhat unrealistically optimistic, believing that the odds of acceptance are (much) greater than the base rate (cf. Weinstein, 1980). We may continue to hold favorable beliefs about the quality of own manuscripts that have been rejected by filtering specific positive information (i.e., typical editorial letters and reviews do contain at least some positive comments) and (to some extent) downplaying the quality of reviews. Also, in evaluating own and others' work, we may tend to assign greater attention and weight to specific domains or criteria that are of special importance to us or that we ourselves excel in (cf. Dunning, Meyerowitz, & Holzberg, 1989). At the same time, the present research cannot rule out one potentially important alternative interpretation. It may be that reviewers submit manuscripts that are actually of greater quality than are the manuscripts that they review. Indeed, not all authors serve as reviewers for peer-reviewed journals, whereas all participants of this study had reviewer experience. As such, the present findings cannot unambiguously be attributed to self-enhancement or self-verification mechanisms.

The major thrust of this study was not its theoretical relevance, but its relevance to understanding judgments of own and others' work in the context of the current peer-review system, an issue that has not yet been truly explored. Indeed, one major question was whether the belief in reviewers' tendencies to SLAM represents an objective reality, a constructed reality, or both. Part of the answer is located in our tendencies to hold unrealistically favorable beliefs regarding our own work, thereby supporting the constructed reality underlying the belief in reviewers' tendencies to SLAM (as well as Levenson's, 1996, sour-grapes hypothesis). At the same time, the current findings suggest (albeit indirectly) that reviewers do tend to recommend rejection on the basis of limitations that would seem to be somewhat less serious in the eyes of editors. Reviewers' appraisals tend to be more negative than those of editors, in that reviewers exhibited greater levels of agreement with editorial decisions favoring rejection than with editorial decisions favoring acceptance. Thus, another reason for why authors sometimes

might be upset is that reviewers factually place particular emphasis on the limitations of a manuscript, thereby supporting the objective reality underlying the belief in reviewers' tendencies to SLAM.

Why exactly might reviewers be overly critical; at least more critical than editors? What are important differences between reviewers and editors? To begin with, reviewers typically tend to have greater expertise regarding the issues addressed in a manuscript than do editors. However, greater expertise and knowledge do not necessarily imply a greater emphasis or recognition of limitations, in that such additional knowledge could also be associated with a greater ability to detect strengths and innovative aspects of a manuscript.

A second explanation is that, given that the base rate for acceptance tends to be quite low within the field of social psychology, reviewers may become inclined to recommend rejection rather than acceptance. For example, in light of this base rate, reviewers may believe that their job is to detect limitations and only occasionally to recommend acceptance. Also, scientists may wish to avoid writing reviews that are expected to deviate significantly from other reviews (cf. Asch, 1955). Given the low base rates for acceptance, the likelihood of being deviant would be greater if one recommends acceptance than if one recommends rejection. Also, it may be psychologically more aversive to deviate by recommending acceptance than by recommending rejection, particularly if reviewers see their role as one of pointing out limitations rather than strengths.

A final explanation may be derived from the fact that in the current peer-review system, editors (unlike reviewers) are accountable. Editors are not anonymous, and editors are the ones whose (negative) decisions might be challenged by the author. Thus, it would seem to be particularly important for editors to carefully attend to both limitations and strengths of a manuscript, whereas reviewers are in a position where they can "almost freely" focus on limitations, rather than strengths.

Turning back to Epstein (1995), it is interesting that one of his major suggestions for improving the review system focuses on increasing reviewer accountability by providing authors with standard forms for evaluating the reviews that they receive. This suggestion is considered to be useful by Fine (1996), who also argues that it would be essential that such forms be completed after a final decision is made on the revised manuscript in order to enhance honest evaluations of the reviews. As pointed out by Fine, Epstein's suggested procedure may (a) elevate review quality; (b) strengthen helpfulness among reviewers; and (c) help authors to feel less dependent on incompetent, biased, or inexperienced reviewers. Moreover, if successful, this procedure may (d) help editors to select capable, constructive, and "balanced" reviewers; and (e) give rise to norms for assessing quality on the basis of limitations and strengths.

Thus, there is good reason to believe that Epstein's (1995) suggested procedure is a useful one (even at the expense of some practical costs). However, this

procedure raises at least two questions. First, it is important to keep in mind that editors need reviewers' honest evaluations and judgments to make a balanced editorial decision. Although enhanced reviewer accountability is likely to result in "more balanced" reviews, it is not clear whether accountability strengthens or diminishes such honesty. Given that this research was anonymous and not relevant to editorial decisions, there are no strong reasons to question the participants' honesty regarding their evaluations of manuscripts that they had reviewed (evaluations which actually were quite critical). Thus, although Epstein's procedure is likely to encourage balanced reviews, the potential risk is that the reviewers tend to be not only less critical but also less honest than they might be when they are less accountable, as in the current peer-review system.

A second issue raised by Epstein's (1995) suggested procedure is whether authors will provide reviewers with fair and useful comments. Although it is possible that such comments are quite fruitful, as suggested by Fine (1996), it is also possible that authors are somewhat unable (or unwilling) to provide "their reviewers" (particularly negative reviewers) with fair and useful feedback. The current findings indicate that authors continue to hold very favorable beliefs about their manuscripts, even after the work has been rejected for publication (i.e., beliefs that were at least as favorable as beliefs about others' work that has been accepted). This finding is understandable, in that (a) it is common for researchers to have considerable faith in their contributions (i.e., based on own logic and own observations); (b) researchers are likely to submit their work to a particular journal, anticipating that this work will be fairly well received (i.e., typically, authors do not submit their work to journals from which they expect a rejection); (c) the base rates for acceptance are low; and, last but not least, (d) it is not easy for many of us to dissociate the quality of our work from our sense of self-esteem. Thus, in light of the finding that authors continue to believe that their rejected manuscripts are quite good, it is—at least to some degree—questionable whether authors are able or willing to provide the reviewer with fair and useful feedback. Similarly, given that reviewers tend to be quite critical of others' work, it is somewhat questionable whether reviewers will consider authors' comments (particularly somewhat critical comments) to be fair and useful.

Clearly, the current research is not without limitations. The sample relied on generosity and interest of the participants, and included only social psychologists with review experience. These unfortunate (yet to some degree inevitable) features limit the scope of this research and make it impossible to rule out alternative interpretations that may be derived from selection bias. Despite these limitations, the findings suggest that, for both good and bad reasons, we as authors may, every now and then, be disturbed about some detailed questions and doubts expressed by reviewers; whereas we as reviewers may often truly believe that such questions and doubts are accurate, appropriate, and useful.

References

Alicke, M., LoSchiavo, F. M., Zerbst, J. I., & Zhang, S. (1997). The person who outperforms me is a genius: Esteem maintenance in upward social comparison. *Journal of Personality and Social Psychology*, **73**, 781-789.

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others. *Social Cognition*, **7**, 275-296.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, **193**, 31-35.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, **57**, 1082-1090.

Epstein, S. (1995). What can be done to improve the journal review process? *American Psychologist*, **50**, 883-885.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, **7**, 117-140.

Fine, M. A. (1996). Reflections on enhancing accountability in the peer review process. *American Psychologist*, **51**, 1190-1191.

Laband, D. N., & Piette, M. J. (1994). A citation analysis of the impact of blinded peer review. *Journal of the American Medical Association*, **272**, 147-149.

Levenson, R. L., Jr. (1996). Enhance the journals, not the review process. *American Psychologist*, **51**, 1191-1193.

Peters, D., & Ceci, S. J. (1985). Peer review: Beauty is in the eye of the beholder. *Behavioral and Brain Sciences*, **8**, 747-750.

Rabinovich, B. A. (1996). A perspective on the journal review process. *American Psychologist*, **51**, 1190.

Reis, H. T., & Stiller, J. (1992). Publication trends in *JPSP*: A three-decade review. *Personality and Social Psychology Bulletin*, **18**, 465-472.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self to be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 209-269). New York, NY: Academic.

Suls, J., & Wills, T. A. (Eds.). (1991). *Social comparison: Contemporary theory and research*. Hillsdale, NJ: Lawrence Erlbaum.

Swann, W. B., Jr. (1983). Self-verification: Bringing social reality in harmony with the self. In J. Suls & A. G. Greenwald (Eds.), *Psychological perspectives of the self* (Vol. 2, pp. 33-66). Hillsdale, NJ: Lawrence Erlbaum.

Swann, W. B., Jr. (1990). To be adored or to be known: The interplay of self-enhancement and self-verification. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 404-448). New York, NY: Guilford.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, **103**, 193-210.

Van Lange, P. A. M., & Rusbult, C. E. (1995). My relationship is better than—and not as bad as—yours is: The perception of superiority in close relationships. *Personality and Social Psychology Bulletin*, **21**, 32-44.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, **39**, 806-820.

Wills, T. A. (1991). Similarity and self-esteem in downward comparison. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 51-78). Hillsdale NJ: Lawrence Erlbaum.